Performance Bounds for Cyberattack Detectors Using Multiple Observations

Onur Toker
Electrical and Computer Engineering
Florida Polytechnic University
Lakeland, FL 33805
otoker@floridapoly.edu

SoutheastCon 2022 | 978-1-6654-0652-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/SoutheastCon48659.2022.9764036

***Abstract*—In this paper, we consider False Data Injection (FDI) attacks with small injected error, and derive risk upper bounds for cyberattack detectors using multiple observations. FDI attacks with small injected error are "slow" attacks which are harder to detect, but such slow attacks can cause major failures if continuously applied over a long time period. A natural question to ask is to what degree the cyberattack detection problem becomes easier if multiple observations acquired over a long time period are used for threat assessment, and the risk level reduction achieved for each new observation. For a cyberattack detector, the false alarm rate is the probability of triggering an alarm when there is no cyberattack, and the probability of miss is the probability of not detecting a cyberattack. The risk level of a cyberattack detector is defined as the sum of the probability of false alarm and the probability of miss. By using the notion of Hellinger distance, we derive bounds on the minimum possible (achievable) risk level under multiple observations, and study asymptotic properties of such bounds. It is proved that that the minimum possible risk level converges to zero exponentially as the number of observations goes to infinity.**

***Index Terms*—Cyberattack Detection, Multiple Observations, Slow Attacks, Risk Bounds, False Data Injection.**

I. Introduction

Cyberattack detection is a very important problem in networked control systems [1]–[3]. Attack detection becomes quite difficult if the difference between the no-attack case and the cyberattack case is quite small. Namely, if the "unusual" behaviour observed in the cyberattack case is very similar to the no-attack case, detection will be difficult. In [4], an edge case with a small injected error is presented to bypass detection and to destabilize the overall system. In other words, "slow" but persistent attacks, i.e. FDI attacks with small injected error, may easily cause a major damage without being detected. For such attacks, it will be more difficult to design a detector which will guarantee a small risk, $P_F + P_M$, where the P_F is the probability of false cyberattack alarm, and P_M is the probability of not detecting the cyberattack. The risk level, $P_F + P_M$, is a quality metric for the detector, hence we consider it as the performance level of the detector. A natural question to ask is the following: What will be the best achievable detector risk level if we try to make a cyberattack assesment by using more and more measurements? How fast the risk level will decrease and converge to zero? A risk level converging to zero very slowly may not be useful from an engineering perspective, because such detectors may detect the attack too late, i.e. well-after a major damage has already been made. In this paper, we provide a theoretical study of this question and illustrate how it can be used for practical cases. We do not design or recommend a particular type of attack detector, instead we study the best achievable risk level among all possible attack detectors and derive performance bounds.

Author would like to acknowledge the support from NSF-1919855, Florida Polytechnic University, and AMI.

Cyberattacks on power systems infrastructure may cause major blackouts, and result equipment and infrastructure damage. Cyberattacks is also an important problem for connected vehicle systems, because such attacks may cause major accidents, loss of life and property. Autonomous vehicles (AV) and advanced driver assistance systems (ADAS) use highly advanced sensors and communication networks which are subject to cyberattacks. A discussion of different attack scenarios for connected vehicles is presented in reports [5], [6]. In [7]–[11], cyberattacks to automotive radars systems are discussed and different attack detection mechanisms are proposed.

A commonly used technique for cyberattack detection is to use an estimator, and generate an alarm if there is a significant difference between estimated and measured values. Basically, a significant difference could be an indication of something unusual. Depending on how this difference value is processed, it is possible to define various types of attack detectors, see [12]–[15], and references therein. Also in [16]–[18], alternative attack detection techniques are explored. In general, the attack detector should be designed to minimize false alarms without significantly degrading the attack detection capability. In other words, it should have a small risk value, $P_F + P_M$, which justifies why the risk level is adopted as a quality metric for cyberattack detectors.

Estimator based techniques are quite useful for detecting cyberattacks at the communication layer, but the accuracy of the estimation model is of crucial for proper operation. In [19], [20], physical-layer attacks on automotive radar sensors are discussed. The first one is based on estimators and the second one is based on a technique called spatio-temporal challenge-response (STCR) which does not depend on an estimation model. A related technique called, physical challange-respone authentication (PyCRA), is introduced in [21]. Both STCR and PyCRA techniques can be quite robust because they do not need an accurate estimator for reliable operation. For physical-

layer attacks, there is a sensor which may or may not be connected to a communication network, and attacks occur at the physical level, i.e. adversarial agents generate physical signals to directly interfere with the sensor's measurement process, confuse the sensor and hence the embedded system processing the sensor output [5], [21]. If the processing nodes, i.e. the embedded systems with sensors, are also communicating with each other, then attacks at the communication layer is also possible. In a good engineering design, both physical-layer and communication layer (if exists) attack detectors should be used to improve safety and reliability.

Our notation is the standard notation used in Measure Theory. This paper is organized as follows: In Section II, we start with some preliminaries needed for the rest of the paper. Main results are presented in Section III, and numerical examples are given in Section IV. Finally, we make some concluding remarks in Section V.

II. Preliminary Results

In this section, we present some preliminary results on matrices, the Hellinger distance, and the total variation norm [22].

A. A bound on the determinant of sum two PD matrices

In this subsection, we prove a generalization of the well-known arithmetic-geometric mean inequality for positive definite (PD) matrices.

Lemma 1 (DSM)**.** *Let A and B two positive definite matrices. If $A \neq B$, then*

$$\det\left(\frac{A+B}{2}\right) > \det(A^{1/2}B^{1/2}) \tag{1}$$

Proof. For 1×1 matrices, this is the usual arithmetic-geometric mean inequality. Therefore, without loss of generality we assume that matrices are of size 2×2 or more. If M is a positive definite matrix, then

$$\det(I+M) > 1 + \det(M).$$

Because, if λ_i, $i = 1, \cdots, n$ are the eigenvalues of M, then $\det(I+M) = \prod_i (1+\lambda_i)$ which is greater than $1 + \prod_i \lambda_i = 1 + \det(M)$. By using this simple inequality, we get

$$\begin{aligned} \det(A+B) &= \det(A)\det(I + A^{-1/2}BA^{-1/2}) \\ &> \det(A)(1 + \det(A^{-1/2}BA^{-1/2})) \\ &> \det(A) + \det(B) \end{aligned}$$

Therefore,

$$\det\left(\frac{A+B}{2}\right) > \frac{\det(A)+\det(B)}{2} \geq \det(A^{1/2}B^{1/2}),$$

where the last inequality follows from the scalar version of the arithmetic-geometric mean inequality. □

B. Hellinger distance and total variation

In this subsection, we define the Hellinger distance, the total variation norm, and prove some inequalities. If $f(x)$ and $g(x)$ are probability density functions defined on $\mathbb{R}^d$, then the Hellinger distance, $H(f,g)$, is defined by

$$\begin{aligned} H^2(f,g) &= \tfrac{1}{2}\int_{\mathbb{R}^d}\left(\sqrt{f(x)} - \sqrt{g(x)}\right)^2 dx \\ &= 1 - \int_{\mathbb{R}^d}\sqrt{f(x)}\sqrt{g(x)}\, dx \end{aligned}$$

and the total variation $TV(f,g)$ is defined as

$$TV(f,g) = \frac{1}{2}\int_{\mathbb{R}^d} |f(x) - g(x)| dx.$$

It is clear that both $TV(f,g)$ and $H(f,g)$ are in $[0,1]$.

Lemma 2 (NPDF)**.** *Let $f(x)$ and $g(x)$ be probability density functions (pdf) defined on $\mathbb{R}^d$. Then,*

$$H^2(f,g) \leq TV(f,g) \leq H(f,g)\sqrt{2 - H^2(f,g)} \leq \sqrt{2}H(f,g).$$

Proof. The first inequality follows from

$$\begin{aligned} 2H^2(f,g) &= \int_{\mathbb{R}^d}(\sqrt{f(x)} - \sqrt{g(x)}\,)^2 dx \\ &\leq \int_{\mathbb{R}^d} |\sqrt{f(x)} - \sqrt{g(x)}\,|\,|\sqrt{f(x)} + \sqrt{g(x)}\,|\, dx \\ &= \int_{\mathbb{R}^d} |f(x) - g(x)| dx = 2TV(f,g) \end{aligned}$$

The second inequality follows from the Cauchy-Schwartz inequality,

$$\begin{aligned} TV^2(f,g) &= (1/4)(\int_{\mathbb{R}^d} |f(x) - g(x)| dx)^2 \\ &\leq (1/4)\int_{\mathbb{R}^d}(\sqrt{f(x)} - \sqrt{g(x)}\,)^2 dx \cdot \\ &\qquad \int_{\mathbb{R}^d}(\sqrt{f(x)} + \sqrt{g(x)}\,)^2 dx \\ &= H^2(f,g)(2 - H^2(f,g)) \end{aligned}$$

where for the last inequality $\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} g(x)dx = 1$ is used. □

C. Hellinger distance for Gaussian distributions

Consider two multivariable Gaussian distributions, $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, and let h be the Hellinger distance between the pdfs. Then

$$h = \left(1 - \frac{\det(\Sigma_1)^{1/4}\det(\Sigma_2)^{1/4}}{\det\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{1/2}} \cdot \exp\left\{-\tfrac{1}{8}(\mu_1-\mu_2)^T\left(\tfrac{\Sigma_1+\Sigma_2}{2}\right)^{-1}(\mu_1-\mu_2)\right\}\right)^{1/2} \tag{2}$$

A proof of this closed form expression is given in [22]. This will be a key result for the rest of the paper.

D. Asymptotic behaviour of the Hellinger distance

For a given pdf, p, defined on $\mathbb{R}^d$, we define

$$p^{(n)}(x_1, \cdots, x_n) = \prod_{k=1}^{n} p(x_k),$$

which will be a pdf defined on $\mathbb{R}^{nd}$. Let f and g be the pdfs for the multivariable Gaussian distributions, $N(\mu_0, \Sigma_0)$ and

$N(\mu_1, \Sigma_1)$, both defined in $\mathbb{R}^d$. In this section, we study the asymptotic behaviour of

$$TV(f^{(n)}, g^{(n)})$$

Both $f^{(n)}$ and $g^{(n)}$ will be multivariable Gaussian distributions. The mean and variance of $f^{(n)}$ and $g^{(n)}$ will be

$$\begin{bmatrix}\mu_1\\ \vdots \\ \mu_1\end{bmatrix}, \begin{bmatrix}\mu_2\\ \vdots \\ \mu_2\end{bmatrix}, \text{ and } \begin{bmatrix}\Sigma_1 & & \\ & \ddots & \\ & & \Sigma_1\end{bmatrix}, \begin{bmatrix}\Sigma_2 & & \\ & \ddots & \\ & & \Sigma_2\end{bmatrix}$$

respectively, where the first two mean values are in $\mathbb{R}^{nd}$ and the last two are block diagonal matrices of size $nd \times nd$.

Let

$$\alpha = \frac{\det(\Sigma_1)^{1/4}\det(\Sigma_2)^{1/4}}{\det\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{1/2}},$$

and

$$\beta = \frac{1}{8}(\mu_1-\mu_2)^T\left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1}(\mu_1-\mu_2).$$

Using the equation (2) for $f^{(n)}$ and $g^{(n)}$ we get

$$H(f^{(n)}, g^{(n)}) = (1-\alpha^n e^{-n\beta})^{1/2}.$$

Using the Lemma 2, we get

$$1-\alpha^n e^{-n\beta} \le TV(f^{(n)}, g^{(n)})$$

and

$$TV(f^{(n)}, g^{(n)}) \le (1-\alpha^n e^{-n\beta})^{1/2}(1+\alpha^n e^{-n\beta})^{1/2}.$$

If f and g are different, then either $\mu_1 \neq \mu_2$ or $\Sigma_1 \neq \Sigma_2$. In the first case, $\beta > 0$ and $\alpha \le 1$ by Lemma 1. In the second case, $\beta \ge 0$ and $\alpha < 1$. These observations can be summarized as follows:

Lemma 3 (ASYM)**.** *Let $f(x)$ and $g(x)$ be pdfs of two different multivariable Gaussian distributions defined on $\mathbb{R}^d$. Then,*

$$\lim_{n\to\infty} TV(f^{(n)}, g^{(n)}) = 1$$

and the convergence is exponential with n. More precisely, the non-negative expression, $1 - TV(f^{(n)}, g^{(n)})$, is bounded from above and below by two functions which converge to 0 exponentially with n.

III. Main Results

Consider a standard binary hypothesis testing problem with pdfs $p_0, p_1 \in L_1(\mathbb{R}^d)$ defined as $p_i(x) = P(x|H_i)$, $i = 0, 1$, and $x \in \mathbb{R}^d$. For a cyberattack detection problem, H_0 may correspond to no attack, and H_1 may correspond to existence of an attack. Assume that we have a detector defined by the complementary regions $D_0, D_1 \subset \mathbb{R}^m$ with $D_0 \cap D_1 = \emptyset$, and $D_0 \cup D_1 = \mathbb{R}^m$. Basically, a cyberattack alarm is triggered iff we have an observation $x \in D_1$, and this is the definition of our cyberattack detector denoted by $A(D_0, D_1)$. The false alarm rate, P_F, and the miss rate, P_M are defined as

$$P_F = P(x \in D_1|H_0),\ P_M = P(x \in D_0|H_1).$$

We define the risk, $P_F + P_M$, as the performance metric of the cyberattack detector $A(D_0, D_1)$, and use the notation $\mathcal{R}A(D_0, D_1)$ to denote this quantity. A detector is considered as "good" iff the risk level, $P_F + P_M$, is "small". For given pdfs $p_0, p_1 \in L_1(\mathbb{R}^d)$, the best possible performance is defined as

$$J(p_0, p_1) = \min_{D_0, D_1} \mathcal{R}A(D_0, D_1)$$

where the minimum is take over all possible complementary regions D_0, D_1. Note that,

$$P_F + P_M = \int_{D_0} p(x|H_1) + \int_{D_1} p(x|H_0),$$

and the minimum will be achieved when the regions D_0 and D_1 are selected as

$$D_0 = \{x : p(x|H_0) > p(x|H_1)\},\ D_1 = \{x : p(x|H_0) \le p(x|H_1)\}.$$

For this specific selection of D_0, D_1

$$\|p_0 - p_1\|_1 = \int_{D_0}(p_0 - p_1) + \int_{D_1}(p_1 - p_0),$$

$$\|p_0 - p_1\|_1 = \int_{D_0} p_0 + \int_{D_1} p_0 + \int_{D_1} p_0 + \int_{D_0} p_1 - \int_{D_1} 2p_0 - \int_{D_0} 2p_1 = 2 - 2J.$$

Therefore

$$J(p_0, p_1) = 1 - \frac{1}{2}\|p_0 - p_1\|_1 = 1 - TV(p_0, p_1).$$

This last equality can be summarized as follows: For given pdfs $p_0, p_1 \in L_1(\mathbb{R}^d)$, there exists cyberattack detectors with small risk level iff the total variation norm, $TV(p_0, p_1)$, is close to one, and the best achievable risk level is equal to $1 - TV(p_0, p_1)$.

A. Multivariable Gaussian case

In this subsection, we study how fast the total variation norm increases as we acquire more and more measurements. If p_0 and p_1 are quite similar, $TV(p_0, p_1)$ will be small and the corresponding $J(p_0, p_1)$ will be close to 1, meaning that minimum possible $P_F + P_M$ will be close 1. In other words, if the pdfs p_0 and p_1 are quite similar, all cyberattack detectors will have poor performance, because the cyberattack detection problem itself will be provably "hard" and it will be impossible to design a cyberattack detector with a "small" risk value. However, as long as p_0 and p_1 are different, no matter how small $TV(p_0, p_1)$ is, by Lemma 3 we know that

$$\lim_{n\to\infty} TV(p_0^{(n)}, p_1^{(n)}) = 1, \text{ and } \lim_{n\to\infty} J(p_0^{(n)}, p_1^{(n)}) = 0$$

and the convergence will be exponential with n. This result can be interpreted as follows: Under the independent observations assumption, even "slow" cyberattack detection problems become exponentially easier with the number of observations.

B. Functional Features

In this subsection, we prove the total variation cannot be increased simply by applying a vector valued function to multidimensional random variables.

Lemma 4 (FF). *Let x and y be $\mathbb{R}^m$ valued random variables with pdfs p and q respectively. Let ψ be a continuous function from $\mathbb{R}^m$ to $\mathbb{R}^r$, and let p_ψ and q_ψ be pdfs of $\mathbb{R}^r$ valued random variables $\psi(x)$ and $\psi(y)$. Then*

$$TV(p_\psi, q_\psi) \leq TV(p, q)$$

Proof. For a given $A \subset \mathbb{R}^r$, we have

$$\begin{aligned} \int_A p_\psi(t)dt &= P(\psi(x) \in A) \\ &= P(x \in \psi^{-1}(A)) = \int_{\psi^{-1}(A)} p(\tau)d\tau. \end{aligned}$$

If s is a ± 1 valued measureable function defined on $\mathbb{R}^r$, we can define

$$A_+ = \{t \in A \,:\, s(A) = +1\}, \quad A_- = \{t \in A \,:\, s(A) = -1\}.$$

It is clear that,

$$\begin{aligned} \int_A p_\psi(t)s(t)dt &= \int_{A_+} p_\psi(t) - \int_{A_-} p_\psi(t)dt \\ &= \int_{\psi^{-1}(A_+)} p(\tau)d\tau - \int_{\psi^{-1}(A_-)} p(\tau)d\tau \end{aligned}$$

hence

$$\int_A p_\psi(t)s(t)dt = \int_{\psi^{-1}(A)} p(\tau)\hat{s}(\tau)d\tau,$$

where $\hat{s}(\tau)$ is another ± 1 valued measureable function defined on $\mathbb{R}^r$. Note that, $\hat{s}(\tau)$ is equal to +1 on $\psi^{-1}(A_+)$ and is equal to -1 on $\psi^{-1}(A_+)$.

As an application of this result, we have

$$\int_A (p_\psi(t) - q_\psi(t))s(t)dt = \int_{\psi^{-1}(A)} (p(\tau) - q(\tau))\hat{s}(\tau)d\tau,$$

and by selecting $s(t)$ as the sign of $p_\psi(t) - q_\psi(t)$ but forcing $s(t)$ to be +1 when $p_\psi(t) = q_\psi(t)$, we get

$$\begin{aligned} \int_A |p_\psi(t) - q_\psi(t)|dt &= \int_{\psi^{-1}(A)} (p(\tau) - q(\tau))\hat{s}(\tau)d\tau \\ &\leq \int_{\psi^{-1}(A)} |p(\tau) - q(\tau)|d\tau. \end{aligned}$$

Now consider the $TV(p_\psi, q_\psi)$,

$$\begin{aligned} 2TV(p_\psi, q_\psi) &= \sup_A \int_A |p_\psi(t) - q_\psi(t)|dt \\ &\leq \sup_A \int_{\psi^{-1}(A)} |p(\tau) - q(\tau)|d\tau \\ &\leq 2TV(p, q), \end{aligned}$$

which completes the proof. □

IV. Numerical Example

In this section, we present a numerical example. For multivariable Gaussian pdfs f and g, we have

$$0.5\ \alpha^{2n}e^{-2n\beta} \leq J(f^{(n)}, g^{(n)}) \leq \alpha^n e^{-n\beta}.$$

where α, β are defined as in Section II.D. Although the best possible cyberattack detector risk level, J, is between these upper and lower limits, for ease of visualization we define a "representative" value as

$$\hat{J}(f^{(n)}, g^{(n)}) = \alpha^{1.5n}e^{-1.5n\beta},$$

which is almost the geometric mean of upper and lower limits. Even though the actual $J(f^{(n)}, g^{(n)})$ and $\hat{J}(f^{(n)}, g^{(n)})$ can be different, this $\hat{J}$ will be quite useful to approximately demonstrate the overlap between two pdfs after n observations.

By using a MATLAB script, we compute $\hat{J}(f^{(n)}, g^{(n)})$ for $n = 1$, $f = N(0, 1)$ and $g = N(\Delta\mu, 1)$, and plot $\hat{J}$ as a function of $\Delta\mu$, see Fig. 1.

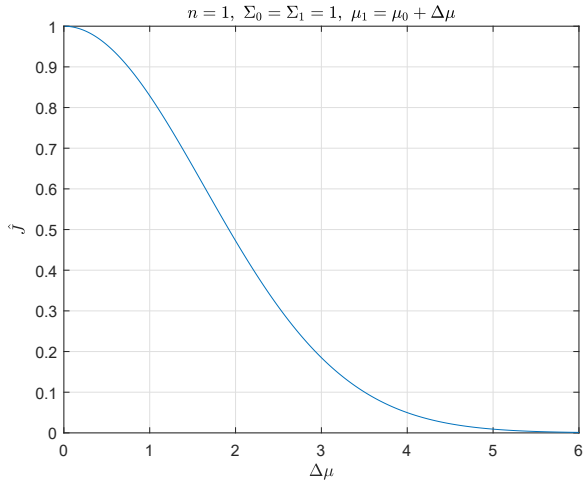

Figure 1. Plot of $\hat{J}$ versus $\Delta\mu$. The vertical axis is a representative value for minimum achievable risk.

We can interpret Fig. 1 as follows: pdfs $N(0, 1)$ and $N(\Delta\mu, 1)$ are difficult to differentiate for small $\Delta\mu$ values, i.e. no matter which cyberattack detector is used risk level will not be small. However, for $\Delta\mu > 5$, these two pdfs are easy to seperate, i.e. it is possible to design a cyberattack detector with a representative risk value less than 0.01.

Consider the Gaussian distributions $f = N(0, 1)$ and $g = N(0.1, 1)$ shown in Fig. 2. These pdfs are very close to each other and difficult to differentiate. Meaning that, no matter which cyberattack detector is used risk level will not be small.

However, if we acquire $n = 2500$ measurements, the pdfs $f^{(n)}$ and $g^{(n)}$ will be quite different with representative risk value of 0.01. Fig. 3 has the equivalent plots of Gaussian distributions $f^{(n)}$ and $g^{(n)}$. Normally, $f^{(n)}$ and $g^{(n)}$ will be $n = 2500$ dimensional pdfs, but the overlap between the two will be similar to the case shown in Fig. 3. This simplification in visualization is possible because of our "representative" risk value definition. Although we have selected the geometric mean of upper and lower limits as the representative risk, other averaging schemes for the representative risk will result similar values.

V. Conclusion

In this paper, we studied the difficulty of detection of false data injection attacks with a very small injected error. Such

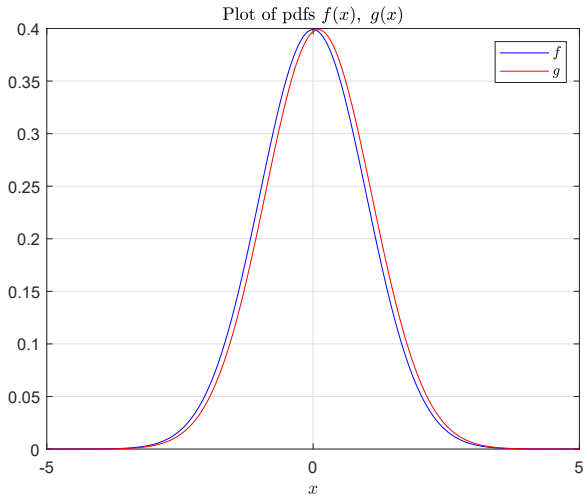

Figure 2. Plots of Gaussian distributions $f = N(0, 1)$ and $g = N(0.1, 1)$.

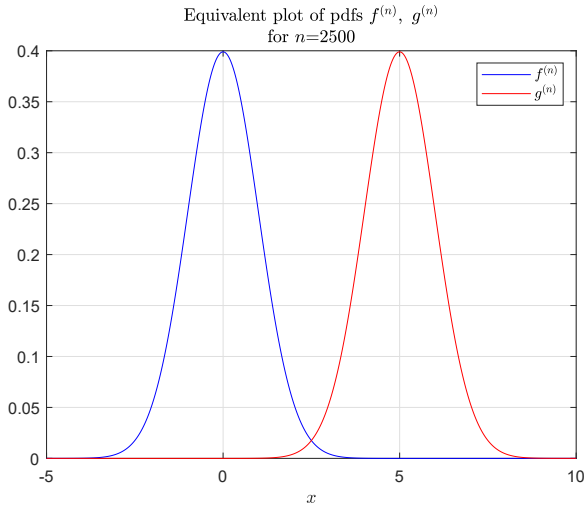

Figure 3. Equivalent plots of Gaussian distributions $f^{(n)}$ and $g^{(n)}$.

small attacks are known to be sufficient for instability if applied persistently over a long period of time. Because of the small injected error, they are hard to detect. However, if the same problem is analyzed with multiple observations, the difficulty, more precisely the overlap between pdfs, gets smaller with increased number of observations. In summary, we have derived upper and lower bounds for the best achievable risk level of a cyberattack detector as a function of the number of measurements, and proved that the minimal risk converges to zero exponentially with the number of measurements.

AKNOWLEDGEMENTS

Funding is provided by NSF-1919855, Advanced Mobility Institute grants GR-2000028, GR-2000029, and Florida Polytechnic University startup grant GR-1900022.

REFERENCES

[1] A. Teixeira, H. Sandberg, and K. H. Johansson, "Networked control systems under cyber attacks with applications to power networks," in *Proceedings of the 2010 American Control Conference*, 2010, pp. 3690–3696.

[2] K. Chatterjee, V. Padmini, and S. A. Khaparde, "Review of cyber attacks on power system operations," in *2017 IEEE Region 10 Symposium (TENSYMP)*, 2017, pp. 1–6.

[3] "Chapter 9 - cybersecurity for the electric power system," in *Cloud Control Systems*, ser. Emerging Methodologies and Applications in Modelling, S. Ison, L. Budd, M. S. Mahmoud, and Y. Xia, Eds. Academic Press, 2020, pp. 271–306. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128187012000172

[4] Y. Mo and B. Sinopoli, "False Data Injection Attacks in Control Systems," in *First Workshop on Secure Control Systems, CPS Week*, Stockholm, Sweden, Apr. 2010.

[5] C. Bhat, "Cybersecurity challenges and pathways in the context of connected vehicle systems," Data-Supported Transportation Operations & Planning Center (D-STOP), Austin, TX, Tech. Rep. 134, Feb. 2018, https://ctr.utexas.edu/wp-content/uploads/134.pdf.

[6] S. Alland, W. Stark, M. Ali, and M. Hegde, "Interference in Automotive Radar Systems: Characteristics, Mitigation Techniques, and Current and Future Research," *IEEE Signal Proc. Mag.*, vol. 36, pp. 45–59, Sep. 2019, https://doi.org/10.1109/MSP.2019.2908214.

[7] O. Toker, S. Alsweiss, J. Vargas, and R. Razdan, "Design of an Automotive Radar Sensor Firmware Resilient to Cyberattacks," in *Proceedings of the 2020 IEEE SoutheastCon*, Raleigh, NC, 2020.

[8] O. Toker and S. Alsweiss, "Design of a Cyberattack Resilient 77 GHz Automotive Radar Sensor," *MDPI, Electronics*, 2020, https://doi.org/10.3390/electronics9040573.

[9] O. Toker, "Physical-layer cyberattack and interference resilient automotive radars," *IEEE Access*, vol. 8, pp. 215 531–215 543, 2020.

[10] O. Toker and B. Kuhn, "A Python Based Testbed for Real-Time Testing and Visualization using TI's 77 GHz Automotive Radars," in *Proceedings of the 2019 IEEE Vehicular Networking Conference*, Los Angeles, CA, 2019.

[11] O. Toker, S. Alsweiss, and M. Abid, "A Computer Vision Based Testbed for 77 GHz mmWave Radar Sensors," in *Proceedings of the 2020 IEEE SoutheastCon*, Raleigh, NC, 2020.

[12] A. Sargolzaei, K. K. Yen, M. N. Abdelghani, S. Sargolzaei, and B. Carbunar, "Resilient design of networked control systems under time delay switch attacks, application in smart grid," *IEEE Access*, vol. 5, pp. 15 901–15 912, 2017.

[13] A. Sargolzaei, K. Yazdani, A. Abbaspour, C. D. Crane III, and W. E. Dixon, "Detection and mitigation of false data injection attacks in networked control systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4281–4292, 2020.

[14] A. Sargolzaei, K. K. Yen, and M. N. Abdelghani, "Preventing time-delay switch attack on load frequency control in distributed power systems," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1176–1185, 2016.

[15] A. Abbaspour, A. Sargolzaei, P. Forouzannezhad, K. K. Yen, and A. I. Sarwat, "Resilient control design for load frequency control system under false data injection attacks," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 9, pp. 7951–7962, 2020.

[16] C. Yang, L. Feng, H. Zhang, S. He, and Z. Shi, "A Novel Data Fusion Algorithm to Combat False Data Injection Attacks in Networked Radar Systems," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 4, pp. 125–136, 2018, https://doi.org/10.1109/TSIPN.2018.2790361.

[17] M. Kordestani, A. Chibakhsh, and M. Saif, "A Control Oriented Cyber-Secure Strategy Based on Multiple Sensor Fusion," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Bari, Italy, Oct. 2019, https://doi.org/10.1109/SMC.2019.8914241.

[18] E. Mousavinejad, X. Ge, Q.-L. Han, F. Yang, and L. Vlacic, "Resilient tracking control of networked control systems under cyber attacks," *IEEE Transactions on Cybernetics*, vol. 51, no. 4, pp. 2107–2119, 2021.

[19] R. G. Dutta, X. Guo, T. Zhang, K. Kwiat, C. Kamhoua, L. Njilla, and Y. Jin, "Estimation of safe sensor measurements of autonomous system under attack," in *Proceedings of the 54th ACM/EDAC/IEEE Design Automation Conference*, Austin, TX, Jun. 2017, https://doi.org/10.1145/3061639.3062241.

[20] P. Kapoor, A. Vora, and K. D. Kang, "Detecting and Mitigating Spoofing Attack against an Automotive Radar," in *Proceedings of the 2018 IEEE Vehicular Technology Conference*, Chicago, IL, Aug. 2018, https://doi.org/10.1109/VTCFall.2018.8690734.

[21] Y. Shoukry, P. Martin, Y. Yona, S. Diggavi, and M. Srivastava, "Attack Resilience and Recovery using Physical Challenge Response Authentication for Active Sensors Under Integrity Attacks," *arXiv:1605.02062v2*, 2016, https://arxiv.org/pdf/1605.02062.

[22] L. Pardo, *Statistical Inference Based on Divergence Measures*, ser. Statistics: A Series of Textbooks and Monographs. CRC Press, 2018. [Online]. Available: https://books.google.com/books?id=ziDGGIkhqlMC